2009-10 Early Warning Indicator Index

Table of Contents

Overview

The Commonwealth's Early Warning Indicator Index (also called the Early Indicator Index, or "Index") uses statewide data from the most recent graduating cohort as the starting point for helping local school districts identify students who may be at risk of not graduating on time from high school. Incoming high school freshmen are assigned one of five risk levels (*Very High Risk*, *High Risk*, *At Risk*, *Borderline*, or *Low Risk*) based on their middle school academic performance and attendance record in school year (SY) 2008-09. Detailed rosters for these students—the majority of whom are likely to be enrolled in grade 9 in SY2009-10—are provided to the Commonwealth's 24 urban school districts[i] in a Microsoft Excel workbook via the Department's online Security Portal.

Thresholds and risk levels were calculated by examining the middle school performance of students in the state's 2008 high school graduation cohort. Using this information, the Massachusetts Department of Elementary and Secondary Education (the "Department") identified those factors which appear to contribute to a student's likelihood of not graduating on time. The 2009-10 Index is based on three indicators:

1. The student's spring 2009 grade 8 MCAS mathematics score;
2. The student's spring 2009 grade 8 MCAS English language arts (ELA) score; and
3. The student's spring 2009 grade 8 attendance rate.

A student's performance with respect to each of these indicators generates "points" which are then added together. A student's risk level will therefore range from *Low Risk* (3-23 points) to *Very High Risk* (51-60 points). Risk levels are assigned to every student for whom the Department has attendance data and at least one year of assessment data.

Research is clear that a variety of factors contribute to a student's likelihood of not graduating on time. Because the Index uses statewide data, school districts may wish to incorporate local "early warning" data to further refine the Index and to inform local analyses of graduation and dropout patterns in the district.

Readers familiar with the history of the Index will note that the SY2008-09 Index incorporated student mobility as a fourth indicator, while the SY2009-10 Index does not. The Department's analyses confirmed that eliminating mobility data from the Index does not substantially weaken the predictability of the results, allowing for an Index that is simpler to assemble and interpret. However, mobility data are included for each student for informational purposes and to aid in the identification of local indicators, if desired.

Version History

The SY2009-10 Index is the continuation of work begun by the Department in cooperation with districts and other partners in May 2008. As of SY2009-10, five versions of the Index have been developed. The first three versions were proof-of-concept models for testing purposes. Rosters based on the latter versions were provided to districts in SY2008-09 and SY2009-10, respectively. Differences among versions comprise the cohort graduation rate data used as the basis of the model, the number of indicators used to predict a student's likelihood of not graduating on time, and the statistical techniques employed to identify thresholds and assign risk levels.

Table 1 – Operational Versions of the Early Indicator Index

	Indicators	Cohort Graduation Rate Data Used to Identify Indicators of On-Time Graduation Probability	Applicable Student Population for Intervention
SY 2008-09	4	2007 Graduating Cohort	2012 Graduating Cohort (Freshmen as of SY2008-09)
SY 2009-10	3	2008 Graduating Cohort	2013 Graduating Cohort (Freshmen as of SY2009-10)

Identification of Indicators of On-Time Graduation Probability

Each year the Department uses the most recent set of cohort graduation data to fit a linear regression model to examine the association between the dependent (outcome) variable (whether the student graduated on time or did not) and a range of independent variables (indicators).

Using data from the 2008 graduation cohort, the Department examined the middle school performance of all students who were expected to graduate from high school in 2008. This process allows the Department to identify a specific set of characteristics that is most likely to predict whether a student who is currently enrolled as a high school freshman is at risk of not graduating on time four years later. The 2009-10 Index examined the association between the outcome variable (where graduating on time = 1 and not graduating on time = 0) and the following 12 indicators:

1. The student's spring 2003 grade 7 MCAS ELA score;
2. The student's spring 2004 grade 8 MCAS mathematics score;
3. The student's spring 2003 grade 7 attendance rate;
4. The student's spring 2004 grade 8 attendance rate;
5. The student's spring 2005 grade 9 attendance rate;
6. The student's age as of September 1, 2008;
7. The student's low income status (i.e., eligibility for free or reduced-price lunch);
8. The student's special education status;
9. The student's English language learner status (i.e., whether the student is Limited English Proficient (LEP));
10. The student's gender;
11. The student's mobility (i.e., number of times the student changed schools, within a district and between districts, while in grades 7-8); and
12. The student's race/ethnicity.

The Department used R, a language and environment for statistical computing, to calculate the Akaike information criterion (AIC) values of regression models employing different sets of indicators. The AIC

values ranged from 12439 (an 11-indicator model, implying the best fit) to 23758 (a 1-indicator model, implying the least fit). In selecting from among the range of models identified, the Department considered the goodness of fit of the model as well as the number and type of indicators. For example, an 11-indicator model is not as user-friendly as a model with fewer indicators; moreover, the Department sought to restrict the model to those indicators within the reasonable control of local educators to provide timely intervention, such as student's middle school attendance and test scores. Although there are several 3- or 4-indicator models with relatively low AIC values (12888 – 14236), they consist of indicators that are beyond the control of local educators (race, gender, low income status or age) or will delay the delivery of interventions (grade 9 attendance). Based on these considerations, the Department selected a model with a relatively low AIC value (14256) consisting of the following three indicators:

1. A student's spring 2009 grade 8 MCAS mathematics score;
2. A student's spring 2009 grade 8 MCAS English language arts (ELA) score; and
3. A student's spring 2009 grade 8 attendance rate.[1]

An important distinction between the SY2009-10 and SY2008-09 versions of the Index is the use of mobility as a fourth indicator. The Department's analyses found that mobility did not have as strong an association with graduating on time as the other indicators, at least as applied to the 2008 graduating cohort. The correlation coefficient was 0.13 for students who did not change school in grades 7 and 8; -.11 for students changing schools once in grades 7 and 8, and -.08 for students changing schools more than once in grades 7 and 8. By contrast, all of the indicators in the three-indicator model had a moderate and positive correlation with the outcome. Mobility was therefore excluded from the model, but information about mobility was included in the student rosters provided to local school districts.

Assigning Thresholds and Index Values

To identify thresholds for assigning risk levels to individual students, the Department examined the performance of students in middle school who were expected to graduate from high school in 2008 by ranking the indicators categorically based on percentile groups. The Department assigned a rank of 20 to cases below the 5th percentile, 19 to cases between the 5th and 10th percentile, 18 to cases between the 10th and 15th percentile, etc., to a rank of 1 to cases above the 95th percentile. The rank percentile value represents the index for each of the three indicators. For example, a student whose Grade 8 MCAS ELA score was above the 95th percentile for ELA scores was assigned an ELA index of 1; a student whose Grade 8 MCAS Math score was between the 5th and 10th percentile for Math scores was assigned a Math index of 19; and a student whose Grade 8 attendance rate between the 90th and 95th percentiles was assigned an attendance index of 2. The higher the index value, the poorer the student's academic performance or attendance, and the greater the risk the student is of not graduating on time as shown below:

[1] A further argument supporting the use of the 3-indicator model is that the model's R^2, or coefficient of determination, is 0.342, meaning that the model explains 34.2% of the variability associated with graduating on time. Based on the standardized coefficients, the probability of graduating on time is expected to increase by 0.13 standard deviations (SD) for each point increase in the grade 8 MCAS ELA scaled score, by 0.1 SD for each point increase in the grade 8 MCAS mathematics scaled score, by 0.1 SD for each point increase in grade 8 attendance, and by 0.32 SD for each point increase in grade 9 attendance after controlling for other indicators. In brief, these indicators show a moderate and positive association with graduating on time; that is, the probability of graduating on time increases as scaled scores and/or attendance rates increase.

Table 2 – Percentiles and Percentile Rank Values

Percentile	Percentile Rank Value	Percentile	Percentile Rank Value
< 5	**20**	50-55	**10**
5-10	**19**	55-60	**9**
10-15	**18**	60-65	**8**
15-20	**17**	65-70	**7**
20-25	**16**	70-75	**6**
25-30	**15**	75-80	**5**
30-35	**14**	80-85	**4**
35-40	**13**	85-90	**3**
40-45	**12**	90-95	**2**
45-50	**11**	> 95	**1**

The sum of the values of the three indicators has a range of 1 - 60. The Department classified these values into 5 percentile groups, ranging from a rank of 10 for cases below the 10th percentile to a rank of 1 for cases above the 90th percentile, as shown in the table below. The table can be interpreted as follows: Of the students belonging to the 2008 graduating cohort classified as "Low Risk", 95 percent graduated on time. Conversely, only 26% of the students classified as "Very High Risk" graduated on time.

Table 3 – 2008 Cohort Graduation Rates by Risk Level

Percentile Rank (Index Value)	Risk Level	Students Included	Graduation Rate
1 – 3 (3 – 23)	**Low Risk**	4,721	95%
4 – 5 (24 - 31)	**Borderline**	2,984	83%
6 – 7 (32 – 40)	**At Risk**	3,314	69%
8 - 9 (41 – 49)	**High Risk**	2,852	48%
10 (50 – 60)	**Very High Risk**	1,662	26%

Handling Partial Data

The Department conducted separate analyses for students with partial data. The Department could have imputed missing values for a student with his or her previous year's test scores. For example, a member of the 2008 graduating cohort with no 2004 grade 8 MCAS mathematics test data could have been assigned their 2002 grade 6 MCAS mathematics data, if available. However, this technique would not have been available to students missing ELA data, because the MCAS ELA test was only administered in grade 7 from 2002 to 2004. Moreover, if students with partial data were included in the overall analysis, the resulting indices for these students would be artificially low.

The Department fitted linear regression models incorporating the following combinations of data sets for students with less than three indicators:

- Grade 7 ELA scaled scores and grade 8 attendance scores;
- Grade 8 mathematics scaled scores and grade 8 attendance scores;
- Grade 7 ELA scaled scores only;

- Grade 8 mathematics scaled scores only; and
- Grade 8 attendance scores only.

The R^2 for each model were 0.114, 0.097, 0.06, 0.06, and 0.06 respectively. Therefore, the Department used the variables contained in the two models with the highest R^2 to predict the probability of graduating in time for students with partial data (the grade 8 attendance rate and either the grade 7 ELA or grade 8 mathematics scaled score). In both cases, attendance data are required for a student to be assigned a risk level.

The sum of the values of the two indicators has a range of 1 - 40. As is the case with the three-indicator model, the Department classified these values into 5 percentile groups, ranging from a rank of 10 for cases below the 10th percentile to a rank of 1 for cases above the 90th percentile as follows:

Table 4 – 2008 Cohort Graduation Rates by Risk Level for Students with Partial Data

Percentile Rank (Index Value)	Risk Level	Students Included	Graduation Rate
1 – 3 (2 – 16)	**Low Risk**	426	73%
4 – 5 (17 - 21)	**Borderline**	278	54%
6 – 8 (22 – 29)	**At Risk**	467	34%
9 (30 – 33)	**High Risk**	123	14%
10 (34 – 40)	**Very High Risk**	157	5%

Generating Student Rosters for the 2013 Cohort

The Department calculated index values and assigned risk levels using 8th grade MCAS performance data and attendance from the 2013 graduating cohort (high school freshmen as of SY2009-10)—the target population for intervention. It is possible, for example, for a student in the 2013 cohort with an identical MCAS and attendance profile as a student belonging to the 2008 cohort to be assigned a different percentile rank and risk level based on the distribution of students in the 2013 cohort as compared to the 2008 cohort. In other words, while 2008 cohort data were used to identify the indicators and set the threshold levels, the MCAS and attendance profiles of students assigned each percentile rank may differ in the 2013 cohort.

Detailed rosters for the 2013 cohort—the majority of whom are likely to be enrolled in grade 9 in SY2009-10—were then distributed to the Commonwealth's 24 urban school districts in a Microsoft Excel workbook via the Department's online Security Portal in the fall of 2009.

Roster Data

The following tables give data definitions found in student rosters, as well as tables showing how Index points and risk levels are calculated. Local school districts may use this information to assign risk levels to students who newly transferred into the district.

Table 5 – Student Roster Information Provided for the 2013 Cohort

Enrollment Data	
SASID	State Assigned Student Identifier
First Name	Student's First Name
Last Name	Student's Last Name
2008-09 School	School enrolled during 2008-09
2007-08 School	School enrolled during 2007-08
2008-09 Grade	Grade (2008-09)
2007-08 Grade	Grade enrolled during 2007-08
Demographic Data	
Age as of 9/1/08	Age as of September 1, 2008
Gender	Gender
Race/Ethnicity	Official Race/Ethnicity (as of June 2009 SIMS[ii])
Years in Mass. Schools	Years enrolled in Massachusetts schools (as of June 2009 SIMS)
Program Data	
Low Income Status	Low Income status (as of June 2009 SIMS)
Limited English Proficient Status	Limited English Proficient status (as of June 2009 SIMS)
Special Education Status	Special Education status[iii] (as of June 2009 SIMS)
Nature of Disability	Nature of student's disability (as of June 2009 SIMS)
Mobility Data	
Mobility Data	Number of times student changed schools between grade 7 and grade 8

Table 6 – Points Assigned Students in the 2013 Cohort (Complete Data)

Grade 8 Mathematics		Grade 8 English Language Arts		Grade 8 Attendance	
Scaled Score	**Points**	**Scaled Score**	**Points**	**Rate**	**Points**
202-208	20	202-216	20	<=79%	20
210-212	19	218	19	80-85%	19
214	18	220-222	18	86-88%	18
--	17	224-228	17	89-90%	17
216	16	230	16	91%	16
--	15	232-234	15	92%	15
218	14	236-238	14	93%	14
--	13	240	13	--	13
220	12	--	12	94%	12
222	11	242	11	95%	11
224-226	10	244	10	96%	10
228-230	9	--	9	--	9
232-234	8	246	8	97%	8
236-238	7	248	7	--	7
240-242	6	250	6	--	6
244-246	5	--	5	98%	5
248-250	4	252-254	4	--	4
252-256	3	256	3	99%	3
258-262	2	258-260	2	--	2
264-280	1	264-280	1	100%	1

Table 7 – Risk Levels Assigned Students in the 2013 Cohort (Complete Data)

Percentile Rank (Index Value)	Risk Level	Students Included
1 – 3 (3 – 23)	**Low Risk**	6258
4 – 5 (24 - 31)	**Borderline**	3762
6 – 7 (32 – 40)	**At Risk**	4221
8 - 9 (41 – 50)	**High Risk**	3982
10 (51 – 60)	**Very High Risk**	1924

Table 8 – Points Assigned Students in the 2013 Cohort with Attendance Data and one MCAS Indicator

Grade 8 Mathematics		Grade 8 English Language Arts		Grade 8 Attendance	
Scaled Score	**Points**	**Scaled Score**	**Points**	**Rate**	**Points**
<=204	20	<=210	20	<=44%	20
206	19	212-214	19	45-57%	19
208	18	216	18	58-65%	18
210	17	218	17	66-73%	17
--	16	--	16	74-81%	16
212	15	--	15	82-85%	15
--	14	220	14	86-88%	14
--	13	--	13	89-90%	13
214	12	--	12	91%	12
216	11	222	11	92%	11
--	10	224	10	93-94%	10
218	9	226	9	95%	9
--	8	228	8	96%	8
220	7	--	7	--	7
222	6	230	6	97%	6
224-228	5	232	5	--	5
230	4	234-238	4	98%	4
234	3	240	3	99%	3
236-242	2	242	2	--	2
>=244	1	>=244	1	100%	1

Table 9 – Risk Levels Assigned Students in the 2013 Cohort with Attendance Data and one MCAS Indicator

Percentile Rank (Index Value)	Risk Level	Student Included
1 – 3 (2 – 16)	**Low Risk**	72
4 – 5 (17 - 21)	**Borderline**	46
6 – 8 (22 – 29)	**At Risk**	69
9 (30 – 33)	**High Risk**	22
10 (34 – 40)	**Very High Risk**	22

[i] 24 urban districts: Boston, Brockton, Cambridge, Chelsea, Chicopee, Everett, Fall River, Fitchburg, Framingham, Haverhill, Holyoke, Lawrence, Leominster, Lowell, Lynn, Malden, New Bedford, Pittsfield, Quincy, Revere, Somerville, Springfield, Taunton, and Worcester.

[ii] Student Information Management System (SIMS): Unique student identifier for all students receiving a publicly funded education in the Commonwealth

[iii] Special Education status: An indication of the educational environment of a student with disabilities, ages 6–21, at the specific time of reporting (e.g., October 1).